Automatic Vehicle Detection in Aerial Videos using SVM

Swaran K Sasidharan, Kishore Kumar N.K

Abstract— The application of image processing techniques in target object detection in aerial videos has become more useful along with the advancement in computer vision applications and increasing need of social security. This paper presents an automatic vehicle detection for aerial videos. Design considers features including vehicle colors and local features, which increases the adaptability and accuracy for detection in various aerial images. The system design consists of background removal and feature extraction stages. Here Gaussian Mixture Models (GMM) is used for background removal, design considers features including color, edge and corner. For vehicle color classification Support Vector Machine (SVM) used. The previous methods for background removal based on histogram approach have the disadvantage of the vehicle pixels being removed if it occurs as a cluster. This drawback is removed in the paper by using a pixel wise classification method called GMM. Afterwards, SVM is used for final classification purpose. Based on the features extracted, a well - trained SVM can find vehicle pixel. Here we also use haar wavelet based feature extraction in post processing stage on detected vehicle it reduces false alarms and enhances the detection rate. This method is applicable for traffic management, military, traffic monitoring etc.

Index Terms—Aerial surveillance, vehicle detection, SVM, Color transform, GMM, Haar, Color space, EM.

————————————— ◆ —————————————

1 INTRODUCTION

A surveillance has wide variety of application in military and civilian uses for monitoring resources such as forests crops and observing enemy activities. Aerial surveillance cover large spatial area and hence it is suitable for monitoring fast moving targets. Vehicle detection in aerial images has important military and civilian uses. It has many applications in the field of traffic monitoring and management.

Detecting vehicle is an important task in areal video analysis. The challenges of vehicle detection in aerial surveillance include camera motions such as panning, tilting, and rotation. In addition, airborne platforms at different heights result in different sizes of target objects. The view of vehicles will vary according to the camera positions, lightning conditions.

Hsu-Yung Cheng [1] utilized a pixel wise classification model for vehicle detection using Dynamic Bayesian Network (DBN). This design escaped from the stereotype and existing frameworks of vehicle detection in aerial surveillance, which are either region based or sliding window based. Hsu-Yung Cheng utilized a histogram removal method for background removal stage in the system design. The histogram approach have the disadvantage that if the number of vehicle or vehicle color is similar to background color this vehicle pixel is removed from the frame during background removal. Stefan Hinz and Albert Baumgartner [2] proposed a hierarchical model that describes the prominent vehicle features on different levels of detail. There is no specific vehicle models assumed, making the method flexible. However, their system would miss vehicles when the contrast is weak or when the influences of neighboring objects are present.

Choi and Yang [3] proposed a vehicle detection algorithm

using the symmetric property of car shapes. However, this cue is prone to false detections such as symmetrical details of buildings or road markings. Therefore, they applied a log-polar histogram shape descriptor to verify the shape of the candidates. Unfortunately, the shape descriptor is obtained from a fixed vehicle model, making the algorithm inflexible. Moreover, similar to [4], the algorithm in [3] relied on mean-shift clustering algorithm for image color segmentation. The major drawback is that a vehicle tends to be separated as many regions since car roofs and windshields usually have different colors. Moreover, nearby vehicles might be clustered as one region if they have similar colors. The high computational complexity of mean-shift segmentation algorithm is another concern.

Lin et al [5] proposed a method by subtracting background colors of each frame and then refined vehicle candidate regions by enforcing size constraints of vehicles. However, they assumed too many parameters such as the largest and smallest sizes of vehicles, and the height and the focus of the airborne camera. Assuming these parameters as known priors might not be realistic in real applications. In [6], the authors proposed a moving-vehicle detection method based on cascade classifiers.

A large number of positive and negative training samples need to be collected for the training purpose. The main disadvantage of this method is that there are a lot of miss detections on rotated vehicles. Such results are not surprising from the experiences of face detection using cascade classifiers. Luo-Wei Tsai and Jun-Wei Hsieh [7] proposed approach for detecting vehicles from images using color and edges. Proposed new color transform model has excellent capabilities to identify vehicle pixels from background even though the pixels are lighted under varying illuminations. Disadvantage this method is it requires large number of positive and negative training samples.

————————————————————

• *Swaran K Sasidharan is currently pursuing master's degree program in Applied Electronics in MG University, India. E-mail: swaranks@gmail.com*
• *Kishore Kumar N.K is currently working as an assistant professor in ICET, Kerala, India. E-mail: kishorekumar439@gmail.com.*
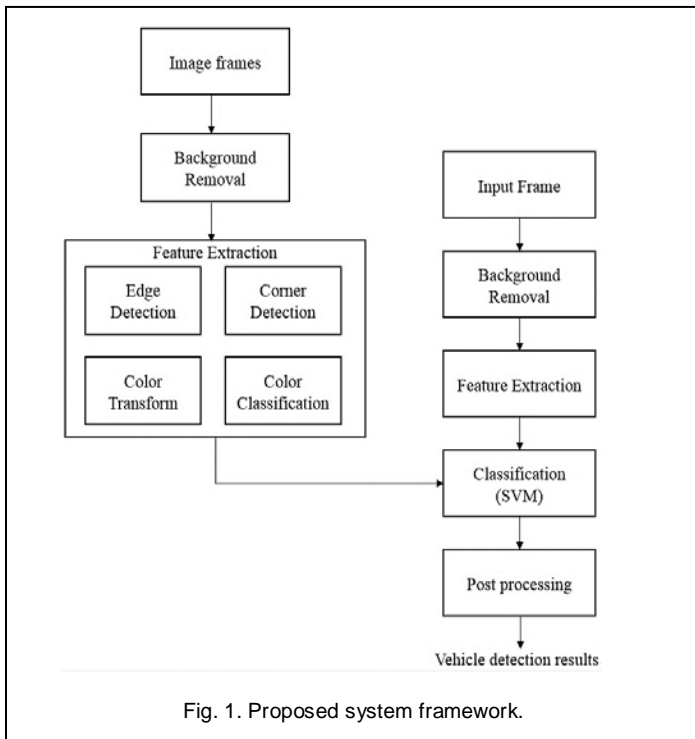
Fig. 1. Proposed system framework.

In the detection phase same feature extraction is also performed as in the training phase. Afterwards the extracted features are used to classify pixels as vehicle pixel or nonvehicle pixel using SVM. In this paper, we do not perform region based classification, which would highly depend on results of color segmentation algorithms such as mean shift. There is no need to generate multi-scale sliding windows either. The distinguishing feature of the proposed framework is that the detection task is based on pixel wise classification. However, the features are extracted in a neighborhood region of each pixel. Therefore, the extracted features comprise not only pixel-level information but also relationship among neighboring pixels in a region. Such design is more effective and efficient than region-based or multi scale sliding window detection methods. The rest of this paper is organized as follows: Section II explains the proposed vehicle detection system in detail. Section III demonstrates and analyses the experimental results. Finally, conclusions are made in Section IV.

2 PROPOSED DETECTION SYSTEM

The proposed system architecture consists of training phase and detection phase. Here we elaborate each block of the proposed system in detail.

2.1 Background Removal

Background removal is often the first step in surveillance applications. It reduces the computation required by the downstream stages of the surveillance pipeline. Background subtraction also reduces the search space in the video frame for the object detection unit by filtering out the uninteresting background.

Here we use Gaussian Mixture Model (GMM) algorithm [9] for background elimination. Flow chat of GMM and EM algorithm is shown in fig .2 and fig .3.

2.2 Feature Extraction

In this stage we extract the local features from the image frame. Here we perform edge detection, corner detection, color transformation and classification.
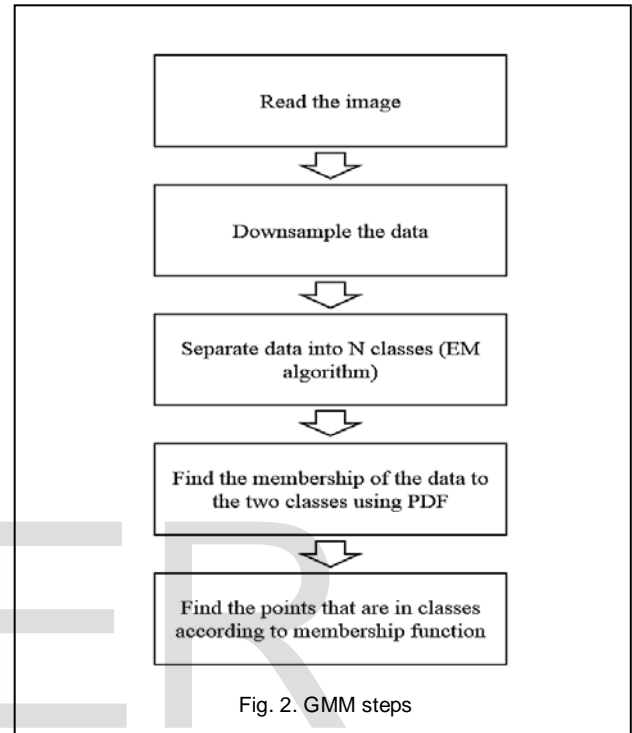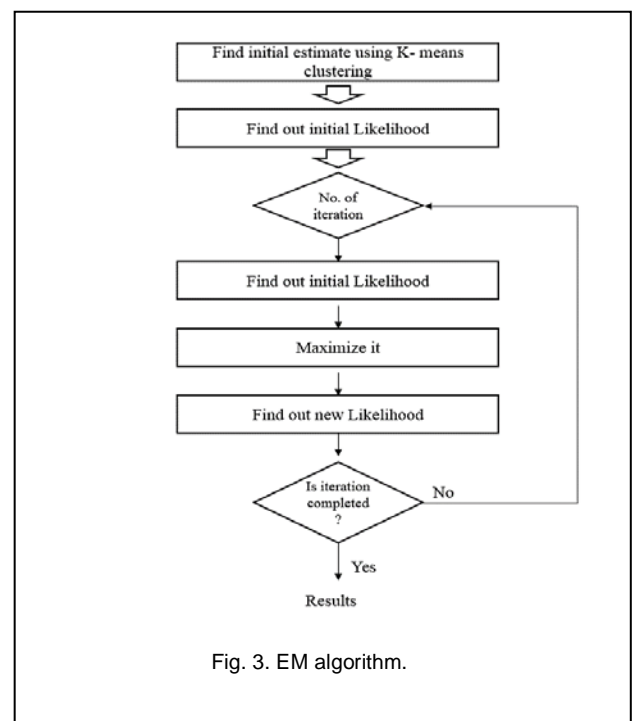

Fig. 2. GMM steps

Fig. 3. EM algorithm.

2.2.1 Edge and Corner Detection

To detect edges we use classical canny edge detector [11]. In canny edge detector there are two thresholds, i.e., the lower threshold T_{low} and higher threshold T_{high}. Then we use Tsai moment-preserving thresholding [12] method to find the thresholds adaptively according to different scenes.

Adaptive thresholds can be found by following derivation consider an image f with n pixels whose gray value at pixel (x,y) is denoted $f(x,y)$. The ith moment m_i of f is defined as

$$m_i = (1/n) \sum_j n_j (z_j)^i = \sum_j p_j (z_j)^i, \quad i = 1,2,3.... \quad (1)$$

Where n_j is the total number of pixel in image f with grey value z_j and $p_j = n_j/n$. For bi-level thresholding, we would like to select threshold T such that the first three moments of image f are preserved in the resulting image g.

$$p_0 z_0^0 + p_1 z_1^0 = m_0, \quad (2)$$

$$p_0 z_0^1 + p_1 z_1^1 = m_1, \quad (3)$$

$$p_0 z_0^2 + p_1 z_1^2 = m_2, \quad (4)$$

$$p_0 z_0^3 + p_1 z_1^3 = m_3, \quad (5)$$

Let all the below –threshold gray values in f be replaced by z_0 and all above- threshold values are replaced by z_1. Then we can solve the equation for p_0 and p_1. After obtaining p_0 and p_1, the adaptive threshold T is computed using

$$p_o = (1/n) \sum_1^T n_j \quad (6)$$

For detecting edges we replace grayscale value $f(x,y)$ with gradient magnitude $G(x,y)$ of each pixel. Adaptive threshold found by equation (6) is used as the higher threshold T_{high} in the canny detector. Then lower threshold is calculated as $T_{low} = 0.1 \times (G_{max} - G_{min}) + G_{min}$, where G_{max} and G_{min} is the maximum and minimum gradient magnitudes in the image.

2.2.2 Vehicle color classification

Here first we transform RGB component of images into u-v plane using color transform then second we use SVM for vehicle color classification.

Color transform and classification using SVM

In [7], the authors introduced a new color transform model that has excellent capabilities to identify vehicle pixels from background. This color model transforms RGB color components into the color domain (u,v), i.e

$$u_p = \frac{2Z_p - G_p - B_p}{Z_p} \quad (7)$$

$$v_p = Max\left\{ \frac{B_p - G_p}{Z_p}, \frac{R_p - B_p}{Z_p} \right\} \quad (8)$$

Where (Rp, Gp, Bp) is the color component of the pixel p and $Zp = (Rp+Gp+Bp)/3$ is used for normalization. It has been shown in [16] that all the vehicle colors are concentrated in a much smaller area on the plane than in other color spaces and are therefore easier to be separated from nonvehicle colors. We can observe that vehicle colors and nonvehicle colors have less overlapping regions under the u-v color model. Therefore, first we apply the color transform to obtain u-v components and then use a support vector machine (SVM) to classify vehicle colors and nonvehicle colors.

As we mentioned on section I, the features are extracted in a neighbourhood region of each pixel. Consider an N × N neighbourhood $\land p$ of pixel p. We extract five types of parameters i.e., S, C, E, A, and Z for the pixel p [1].
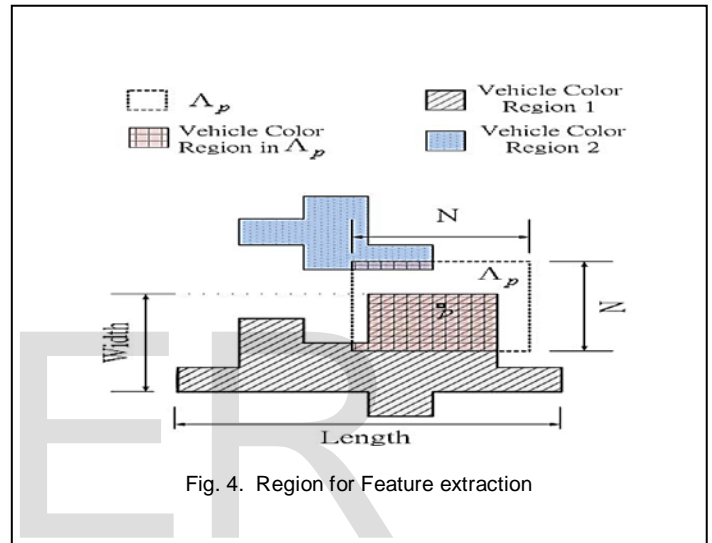

Fig. 4. Region for Feature extraction

The first feature S denotes the percentage of pixels in $\land_p$ that are classified as vehicle colors by SVM. $N_{vehicle\ color}$ denotes the number of pixels in $\land p$ that are classified as vehicle colors by SVM. Similarly, N_{Corner} denotes to the number of pixels in $\land p$ that are detected as corners by the Harris corner detector, and N_{Edge} denotes the number of pixels in $\land_p$ that are detected as edges by the enhanced Canny edge detector

$$S = \frac{N_{Vehicle\ color}}{N^2} \quad (9)$$

$$C = \frac{N_{corner}}{N^2} \quad (10)$$

$$E = \frac{N_{Edge}}{N^2} \quad (11)$$

The last two features and are defined as the aspect ratio and the size of the connected vehicle-color region where the pixel resides, as illustrated in Fig.4. More specifically A= length/Width, and feature Z is the pixel count of "vehicle color region 1" in Fig. 4.

2.3 Vehicle Classification by SVM

Fig. 5. Classification using SVM.
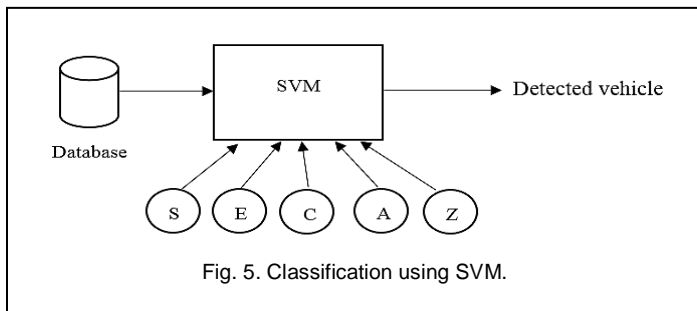
An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on. Here we use the extracted features such as S, E, C, A and Z for training Support Vector Machine (SVM), during this training phase we create a database of vehicle features. After training phase the stored database is used for detection purpose.

2.4 Post Processing

In post processing stage we enhance the detection and performs the labelling of connected components to get vehicle objects. Size and aspect ratio constraints are applied again after morphological operations in the post processing to eliminate unwanted objects i.e., nonvehicle objects.

Here we use haar wavelet based future extraction [17] to enhance the detection on vehicles classified by SVM. In all cases, classification is performed using GMM. Wavelets are essentially a type of multi resolution function approximation that allow for the hierarchical decomposition of a signal or image. They have been applied successfully to various problems including object detection [18] [19], face recognition [20] and image retrieval [21]. Different reasons make the features extracted using Haar wavelets attractive for vehicle detection. Some of them are they form a compact representation, they encode edge information which is an important feature for vehicle detection, they capture information from multiple resolution levels and also there exist fast algorithms for computing these features [19] [18]. Several reasons make these features attractive for vehicle detection. First, they form a compact representation. Second, they encode edge information, an important feature for vehicle detection. Third, they capture information from multiple resolution levels. Finally, there exist fast algorithms, especially in the case of Haar wavelets, for computing these features.

The input image is resized to 64 x 64 image which is decomposed into 2 levels. In the training stage we calculate sigma, variance and mean of wavelet coefficients and classify it into two group's vehicle and non-vehicle. In detection stage four coefficients are fed into the multi Gaussian models it is then compared with the trained data and found out the maximum value of mean. From the calculated value of mean vehicle pixel is identified.

3 EXPERIMENTAL RESULTS

Fig. 6. Snapshots of the experimental videos.
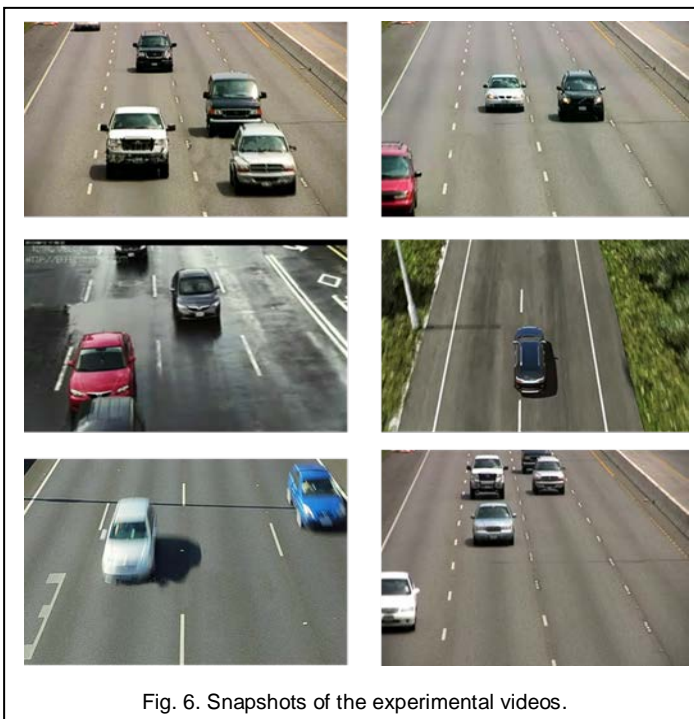
To analyze the performance of the proposed system, image frames from various video sequences with different scenes and different filming altitudes are selected. The experimental images are displayed in Fig. 6.

3.1 Background Removal Results

We use GMM for background removal which is more efficient than the existing histogram based method. In histogram based the input image histogram bin is quantized into 16x16x16. Colors corresponding to the first eight highest bins are regarded as background colors and removed from the scene. Fig. 7 displays an original image frame, and Fig.8 display the corresponding image after background removal.

The histogram approach has the disadvantage that if the number of vehicle with same color is more or vehicle color is similar to background color then vehicle pixel is removed from the frame. This drawback is eliminated by using GMM. In GMM method each image pixel is classified into foreground or background according a membership function.
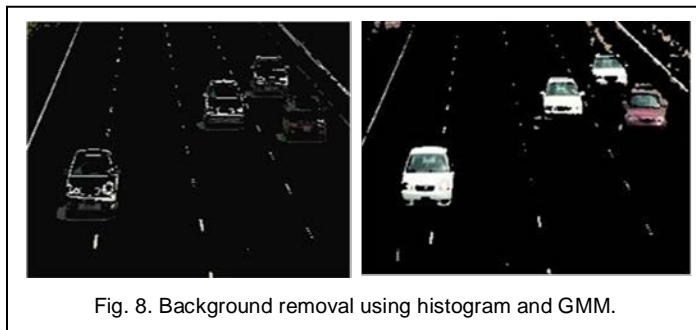

Fig. 7. Input frame.
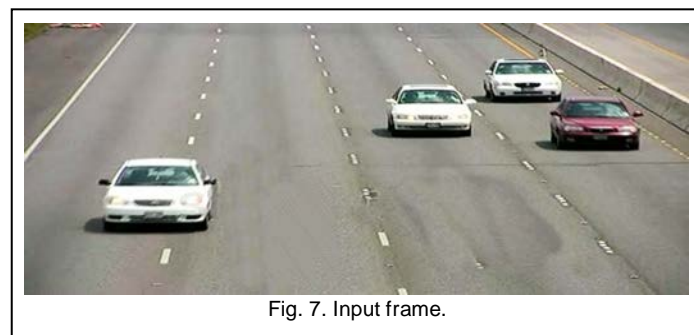
Fig. 8. Background removal using histogram and GMM.

3.2 Color Classification Results

In this stage vehicle and nonvehicle objects are classified according to the color they have. In classification vehicles are represented by binary 1 and nonvehicles are represented by binary 0. When using SVM, we need to select the block size to form sample.
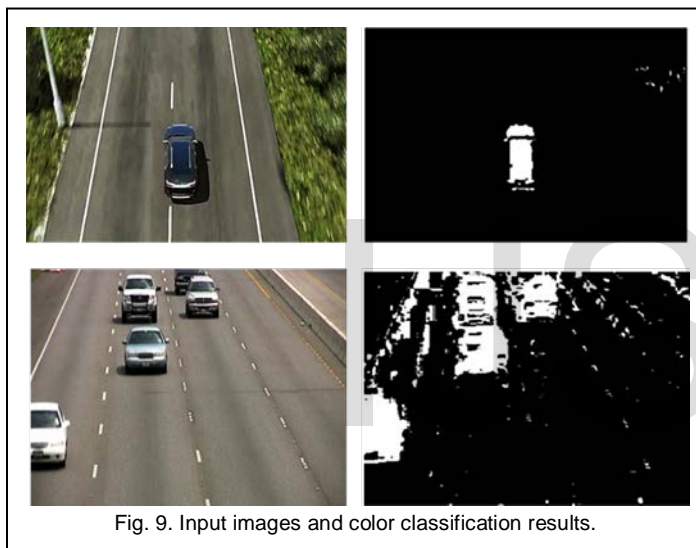

Fig. 9. Input images and color classification results.

According to the observation, we take each 3 x 4 block to form a feature vector. The color of each pixel would be transformed to u and v color components using (7) and (8). The training images used to train the SVM are displayed in Fig. 4.3. Notice that the blocks that do not contain any local features are taken as nonvehicle areas without the need of performing classification via SVM. Fig. 9 shows the results of color classification by SVM after background color removal and local feature analysis.

3.3 Vehicle Detection Results

Here we use Support Vector Machine (SVM) for final vehicle classification. The extracted parameters ie, S, E, C, A, Z are used to train SVM. We can observe that the neighbourhood area ∧p with the size of 7x7 yields the best detection accuracy. Therefore, for the rest of the experiments, the size of the neighbourhood area for extracting observations is set as 7x7.

We compare different vehicle detection methods. Vehicle detection method proposed in [1] produces lot of false detection if same colored vehicles in the frame are high also
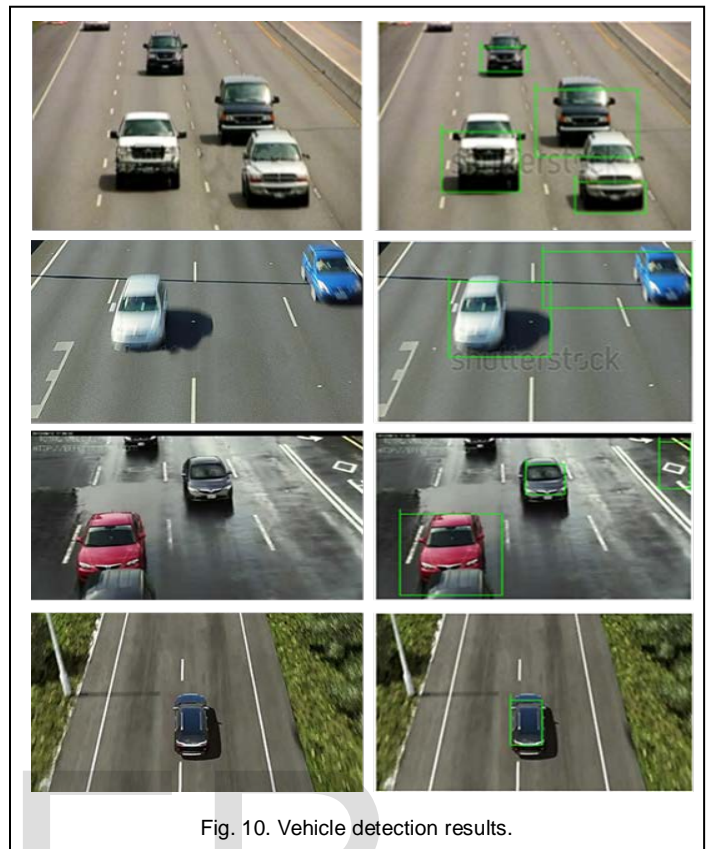

Fig. 10. Vehicle detection results.

rectangular structures in the image frame are detected as vehicles. The moving-vehicle detection with road detection method in [14] requires setting a lot of parameters to enforce the size constraints in order to false detections. However, for the experimental data set, it is difficult to select one set of parameters that suits all videos. Setting the parameters heuristically for the data set would result in low hit rate and high false positive numbers. The cascade classifiers used in [15] need to be trained by a large number of positive and negative training samples. The number of training samples required in [15] is much larger than the training samples used to train the SVM classifier. The colors of the vehicles would not dramatically change due to the influence of the camera angles and heights. However, the entire appearance of the vehicle templates would vary a lot under different heights and camera angles. When training the cascade classifiers, the large variance in the appearance of the positive templates would decrease the hit rate and increase the number of false positives. Moreover, if the aspect ratio of the multiscale detection windows is fixed, large and rotated vehicles would be often missed. The symmetric property method proposed in [16] is prone to false detections such as symmetrical details of buildings or road markings. Moreover, the shape descriptor used to verify the shape of the candidates is obtained from a fixed vehicle model and is therefore not flexible. Moreover, in some of our experimental data, the vehicles are not completely symmetric due to the angle of the camera. Therefore, the method in [16] is not able to yield satisfactory results.

Compared with these methods, the proposed method does not depend on strict vehicle size or aspect ratio constraints. The results demonstrate flexibility and good generalization ability on a wide variety of aerial surveillance scenes under different heights and camera angles.

4 CONCLUSION

In this paper, we have proposed an automatic vehicle detection system for aerial images that does not depends on camera heights, vehicle sizes, and aspect ratios. In this system, we have not performed region-based classification, which would highly depend on computational intensive color segmentation algorithms such as mean shift. The proposed detection system uses pixel wise classifications. Proposed detection system uses Gaussian Mixture Models (GMM) for background removal which is more efficient than the existing histogram based methods. Here we also use haar wavelet based feature extraction in post processing stage to reduce false detections.Which enhances the detection rate. The experimental results demonstrate flexibility and good generalization abilities of the proposed method on a challenging data set with aerial surveillance images taken at different heights and under different camera angles.

REFERENCES

[1] Hsu-Yung Cheng, Chih-Chia Weng, and Yi-Ying Chen, "Vehicle Detection in Aerial Surveillance Using Dynamic Bayesian Networks", IEEE Trans. Image Process., vol. 21, no. 4, pp. 2152–2159, Apr. 2012.

[2] S. Hinz and A. Baumgartner, "Vehicle detection in aerial images using generic features, grouping, and context," in Proc. DAGM-Symp., Sep. 2001, vol. 2191, Lecture Notes in Computer Science, pp. 45–52.

[3] J. Y. Choi and Y. K. Yang, "Vehicle detection from aerial images using local shape information," Adv. Image Video Technol., vol. 5414, Lecture Notes in Computer Science, pp. 227–236, Jan. 2009.

[4] H. Cheng and D. Butler, "Segmentation of aerial surveillance video using a mixture of experts," in Proc. IEEE Digit. Imaging Comput.—Tech. Appl., 2005, p. 66

[5] R. Lin, X. Cao, Y. Xu, C.Wu, and H. Qiao, "Airborne moving vehicle detection for urban traffic surveillance," in Proc. 11th Int. IEEE Conf. Intell. Transp. Syst., Oct. 2008, pp. 163–167.

[6] R. Lin, X. Cao, Y. Xu, C.Wu, and H. Qiao, "Airborne moving vehicle detection for video surveillance of urban traffic," in Proc. IEEE Intell. Veh. Symp, 2009, pp. 203–208.89.

[7] L. W. Tsai, J. W. Hsieh, and K. C. Fan, "Vehicle detection using normalized color and edge map," IEEE Trans. Image Process., vol. 16, no. 3, pp. 850–864, Mar. 2007.

[8] N. Cristianini and J. Shawe-Taylor, An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods, Cambridge, U.K.: Cambridge Univ. Press, 2000.

[9] Douglas Reynolds, Gaussian Mixture Models, MIT Lincoln Laboratory, 244 Wood St., Lexington, MA 02140, USA.

[10] Pushkar Gorur, Bharadwaj Amrutur, "Speeded up Gaussian Mixture Model Algorithm for Background Subtraction", in Proc. 8th Int. IEEE Conf. Advanced Video and Signal-based Surveillance.,2011, pp. 386–391.

[11] J. F. Canny, "A computational approach to edge detection," IEEE Trans. Pattern Anal. Mach. Intell., vol. PAMI-8, no. 6, pp. 679–698, Nov. 1986.

[12] W. H. Tsai, "Moment-preserving thresholding: A new approach," Comput. Vis.Graph. Image Process, vol. 29, no. 3, pp. 377–393, 1985.

[13] Sean Borman, The Expectation Maximization Algorithm A short tutorial.

[14] A. C. Shastry and R. A. Schowengerdt, "Airborne video registration and traffic-flow parameter estimation," IEEE Trans. Intell. Transp. Syst., vol. 6, no. 4, pp. 391–405, Dec. 2005.

[15] H. Cheng and J.Wus, "Adaptive region of interest estimation for aerial surveillance video," in Proc. IEEE Int. Conf. Image Process., 2005, vol. 3, pp. 860–863.

[16] L. D. Chou, J. Y. Yang, Y. C. Hsieh, D. C. Chang, and C. F. Tung, "Intersection- based routing protocol for VANETs,"Wirel. Pers. Commun., vol. 60, no. 1, pp. 105–124, Sep. 2011.

[17] Zehang Sun, George Bebis and Ronald Miller, "Quantized Wavelet Features and Support Vector Machines for On-Road Vehicle Detection," Computer Vision Laboratory, Department of Computer Science, University of Nevada.

[18] C. Papageorgiou and T. Poggio, "A trainable system for object detection," International Journal of Computer Vision, vol. 38, no. 1, pp. 15-33, 2000.

[19] H. Schneiderman, A statistical approach to 3D object detection applied to faces and cars. CMU-RI-TR-00- 06, 2000.

[20] G. Garcia, G. Zikos, and G. Tziritas, "Wavelet packet analysis for face recognition," Image and Vision Computing, vol. 18, pp. 289-297, 2000.

[21] C. Jacobs, A. Finkelstein and D. Salesin, "Fast multiresolution image querying," Proceedings of SIGGRAPH, pp. 277-286, 1995.

[22] Guo, D., Fraichard, T., Xie, M., Laugier, "Color Modeling by Spherical Influence Field in Sensing Driving Environment", IEEE Intelligent Vehicle Symp, pp. 249-254, 2000.

[23] Mori, H., Charkai, "Shadow and Rhythm as Sign Patterns of Obstacle Detection," Proc. Int'l Symp. Industrial Electronics, pp. 271-277, 1993.

[24] Hoffmann, C., Dang, T., Stiller, "Vehicle detection fusing 2D visual features," IEEE Intelligent Vehicles Symposium, 2004.

[25] Kim, S., Kim, K et al, "Front and Rear Vehicle Detection and Tracking in the Day and Night Times using Vision and Sonar Sensor Fusion, Intelligent Robots and Systems," IEEE/RSJ International Conference, pp. 2173-2178, 2005.